Slide 1 of 14

Title Slide: Sample Designs, Weights, and Variances Across the Secondary Longitudinal Studies from 1972-2000

Slide 2 of 14

As discussed in the common modules, when analyzing data from complex sample surveys, certain procedures must be used to assure that estimates made from the data are representative of the population and that hypothesis tests are accurate. This module provides information about the weights that must be applied within the National Longitudinal Study of 1972 (or NLS-72), High School and Beyond (or HS&B), and the National Education Longitudinal Study of 1988 (or NELS:88) to ensure that estimates are representative of the specific study population. This module will also describe the types of weights within NLS-72, HS&B, and NELS:88 and provide guidance regarding how to select the appropriate weight for specific analyses. Additionally, this module will describe appropriate procedures for calculating standard errors using NLS-72, HS&B, and NELS:88 data. For information about weighting and calculating appropriate standard errors more generally, please view the relevant common modules, if you have not already done so, which are accessible by clicking on the underlined screen text on this slide.

Slide 3 of 14

The first step to understanding NLS-72, HS&B, and NELS:88 sampling designs and associated study weights is acknowledging that there are not any short cuts that can be taken when working with these studies from the past. Even though there are major design features common across NLS-72, HS&B, and NELS:88, each of these data sets contain unique variations in methodologies used that require analysts to become familiar with the technical documentation and reports that have been written. As a variety of contractors worked on these datasets, there is not always consistency in how the studies were designed, how the data files were set up, and how the weights were created. In order to fully understand the sampling designs and associated weights, you MUST review each study's documentation and understand how the data were collected, coded, and weighted to ensure accurate analyses.

Slide 4 of 14

NLS-72, and all NCES secondary longitudinal studies, use a two-stage sample design. In the first stage, high schools were selected, and then in the second stage, students within those schools were selected.

The target population for NLS-72 was all public and private schools in the 50 states and the District of Columbia that enrolled twelfth graders during the 1971-72 school year. The study excluded students from schools for the physically or mentally handicapped, those for legally confined students, and in special situations (such as vocational schools) where students were also enrolled in other high schools in the sampling frame.

In the first stage of sampling, schools were selected from the computerized school files maintained by the Office of Education and the National Catholic Education Association with a probability of selection proportional to the estimated number of senior students for the smallest strata (fewer than 300 seniors) and with equal probabilities for the remaining size strata. Next, schools were stratified by type of control (public or nonpublic), geographic region (Northeast, North Central, South, and West), grade 12 enrollment (three size categories), proximity to institutions of higher learning (3 distance categories), percent minority group enrollment (eight categories, public schools only), income level of the community (11 categories for public schools and eight categories for Catholic schools), and degree of urbanization (10 categories). Schools in low-income areas and schools with a high proportion of minority group enrollment were oversampled to ensure there would be enough schools in the sample to allow researchers to generate unbiased estimates.

From this stage of sampling, 1,200 eligible schools with a 12th grade were selected, and 1,061 schools responded.

Slide 5 of 14

In the second stage of sampling, a fixed number of about 18 students per school were randomly selected from rosters provided by the schools. Students were selected from 12th graders within the school, not by classroom.

The result of the sampling process was a sample of 16,683 12th graders representing 2,950,000 students. The supplementary sample from the first-follow up added 4,450 more 1972 high school seniors from 257 additional schools to the nationally representative base year sample.

Slide 6 of 14

HS&B, like NLS-72, used a two-stage sample design. In the first stage, high schools were selected, and then in the second stage, students within those schools were selected.

In HS&B, the target population was public and private high schools in the 50 states and the District of Columbia. In the first stage of sampling, schools were selected from the frame of the Common Core of Data (CCD) and the School Universe Computer file distributed by the Curriculum Information Center, Inc. (CIC) of Denver, Colorado. Public schools with high percentages of Hispanic students, Catholic schools with high percentages of minority students, alternative public schools, and private schools with high-achieving students were oversampled to ensure there would be enough schools in the sample to allow researchers to generate unbiased estimates.

The study excluded students from schools for the physically or mentally handicapped, those for legally confined students, and in special situations (such as vocational schools) where students were also enrolled in other high schools in the sampling frame.

Slide 7 of 14

In the second stage of sampling, a fixed number of 36 sophomores and 36 seniors from each of the 1,015 schools were randomly selected from rosters provided by the schools. Students were randomly selected within the school, not by classroom, meaning the HS&B sample is clustered. Special efforts were made to identify sampled students who were twins or triplets so that their co-twins or co-triplets could be invited to participate in the study.

The result of the sampling process was a sample of 30,030 sophomores, representing 3.7 million 1980 high school sophomores, and 28,240 seniors representing 3.0 million 1980 high school seniors.

Slide 8 of 14

NELS:88 used a two-stage sample design. In the first stage 40,000 public and private schools serving approximately 3,000,000 8th graders in the 50 states and the District of Columbia were identified using the Common Core of Data (or CCD) and the Quality Education Data (or QED).

Next, schools were stratified by region (Northeast, South, North Central, and West), urbanicity (urban, suburban, and rural), and school type (public, Catholic, and other private). Asian/Pacific Islander and Hispanic students, as well as private schools, were oversampled to ensure there would be enough students in the sample to allow researchers to generate unbiased estimates. From this stage of sampling, 1,052 schools provided base year data.

Slide 9 of 14

In the second stage of sampling, 24-26 eighth grade students were randomly selected from rosters provided by the schools. These students became the core sample for the school. After the initial 24-26 8th grade students were selected, Asian/Pacific Islander and Hispanic students were identified from rosters of non-selected students. Between one to two Asian/Pacific Islander and Hispanic students were then selected and added to the sample. It is important to note that the NELS:88 sample excluded approximately 5% of students listed on school rosters because school personnel indicated that students with mental disabilities, physical disabilities, and language difficulties should not participate in the study.

Like the other secondary longitudinal studies, students were selected within the school meaning the NELS:88 sample is clustered.

The result of the sampling process was a sample of 24,599 8th grade students representing approximately 3 million students.

Slide 10 of 14

As was discussed in the common module titled, 'Statistical Analysis of NCES Datasets Employing a Complex Design', weights must be used in analyses of data from studies like NLS-72, HS&B, and NELS:88 in order to make estimates produced from the sample representative of the target population.

As NLS-72, HS&B, and NELS:88 are not simple random samples, weights account for probability of selection and are an indicator of the relative importance of a particular observation to the computation of a sample statistic. Additionally, not everyone sampled responded to NLS-72, HS&B, and NELS:88. Therefore, weights also account for differential response patterns within a given round (also referred to as wave or sometimes follow-up), as well as across time, to ensure that the data are representative of the population even though some members of the sample do not respond to the survey.

Ideally, there would be a nonresponse adjusted weight available for every component of every round of data collection. However, in longitudinal studies that have multiple components across multiple rounds of data collection, there are a number of possible weights that can be created for analysis of data within and across rounds. It is not economical, nor useful in a practical sense, to create nonresponse adjusted weights for every combination of components across every round of data collection. Therefore, each researcher will need to decide which of the weights provided is the best one to use, given a specific research question.

Slide 11 of 14

Even though the Secondary Longitudinal Studies from 1972-2000 (NLS-72, HS&B, and NELS:88) are similar, it is important to read the documentation regarding sampling and weights for the specific data set of interest.

For example, when working with data from all six rounds of NLS-72, there is one set of weights that must be used when working with data from the base year through the fourth follow-up (22,652 cases), and another set of weights that must be used when working with the fifth follow-up (14,489 cases).

Within the HS&B sophomore cohort file, there is a base year weight for the entire 30,030 sample members as of 1980. At the same time, there is a separate base year weight for the 14,825 cases that are part of the fifth follow-up sample. Though both weights are available, users should use the weight that reflects the sophomores that are part of the fifth follow-up. This weight is called BYWT. It is important to note that for HS&B, reduced data sets are provided in the current data releases (containing 14,825 sophomores and 11,995 seniors).

Slide 12 of 14

Within NELS:88, there is only one base year cross-sectional weight BYQWT. This weight is only provided on the NELS:88/92 data file; it cannot be accessed via the NELS:88/94 or NELS:88/2000 data files.

The only way to learn how to appropriately use weights within the Secondary Longitudinal Studies from 1972-2000 is to read the documentation. To facilitate your understanding, a resource document has been provided that includes additional detail regarding the weights and extant data files that are available for each of the secondary longitudinal studies from 1972-2000. This resource document can be accessed by clicking the underlined screen text, 'appropriately use weights'.

Slide 13 of 14

In the common module titled, 'Statistical Analysis of NCES Datasets Employing a Complex Sample Design', two standard error calculation procedures were discussed: Replication Techniques and Taylor Series linearization. Replication is a method that calculates appropriate standard errors based on differences between estimates from the full sample and a series of created subsamples, or replicates. The Taylor series linearization method uses PSU and strata identifiers to compute the appropriate standard errors.

Analysts using NLS-72, HS&B, and NELS:88 data will need to consult the technical documentation associated with each study to determine which method should be used to compute standard errors. Documentation for each study can be accessed by clicking the corresponding underlined screen text.

Slide 14 of 14

This module provided information about the weights that must be applied within NLS-72, HS&B, and NELS:88 to ensure that estimates are representative of the specific study population. This module also described the types of weights within NLS-72, HS&B, and NELS:88 and provided guidance regarding how to select the appropriate weight for specific analyses. Additionally, this module described appropriate procedures for calculating standard errors using NLS-72, HS&B, and NELS:88 data.

You may now proceed to the next module in the series, or click the exit button to return to the landing page.